

Note on program:

- to run tests with statistics enter the following command
`py.test --capture=no --hypothesis-show-statistics --debug propertytests.py`
- python has much more restrictions for element generation and I could not find the ability to adjust a test to pass for a small percentage of fails. For the group fairness tests, I needed to generate multiple rows of inputs instead of just one row per test run in order to calculate statistics to determine if the test passed. Therefore, I opted not to use strategies for the group fairness tests and just generate rows by calling a method in the test.

3. Report results

There only test that passed is Sex Group Fairness. Although there is group fairness for sex, there is no causal fairness for sex indicating some discrimination. Age did not pass for both group and causal fairness indicating clear discrimination for age.

4.

The problem could have been identified by considering the efficacy of the data and if the data itself was bias. It is also the data engineers responsibility to ask themselves what variables in the data would be unethical if they were used to predict an outcome in the model. For this data set, one can easily argue the it is unethical to grant credit approval based on someone's sex or age. To consider whether groupings are ethical additional people to a software engineer should participate in the discussion. I believe, economists/social scientist, statisticians, and experts in the field of study of the data should be consulted. Social Scientist/Economist have a better understanding of human behavior and specifically study predictors of behaviors and statisticians know how to best interpret data where someone else might make false correlations. Lastly, an expert in the field of the data's study will add more context to data groups. An example is a doctor noting when race and gender is an ethical grouping to a study. Race is a relevant grouping for genetic diseases but it's absolutely irrelevant to medicine dosing.

5. Fairness testing is closed-box testing. Fairness deals with groupings of *inputs* and how they correlate with an *output*, which is part of the definition of closed-box testing, checking for preconditions and postconditions. Fairness testing is not concerned with how the model predicts the outcome but rather that the outcomes are fair.

6. I would simply just write tests on other properties of the system, such as general input handling, precondition failures, post conditions failure etc. This way I can ensure

that general usability of the system works. Then I could go back and improve the training model and run property tests.